



Ljudmila PETKOVIĆ

32 boulevard du maréchal Joffre  
92340 Bourg-la-Reine, France

✉ +33 (0)7 68 81 83 47

✉ [ljudmila.petkovic@sorbonne-universite.fr](mailto:ljudmila.petkovic@sorbonne-universite.fr)

ID 0000-0003-2694-4096

## Intérêts scientifiques

Mes recherches portent sur les humanités numériques et le traitement automatique des langues. Je suis particulièrement intéressée par les méthodes de diffusion, de partage et de valorisation des fonds patrimoniaux au prisme des humanités numériques en vue du repérage des circulations des savoirs. **Humanités numériques** : reconnaissance optique de caractères (OCR), reconnaissance de textes manuscrits (HTR), structuration des données en XML-TEI, publication des éditions numériques interrogeables.

**Traitement automatique des langues** : fouille de textes, linguistique computationnelle.

## Formation

2021 – **Doctorat en humanités numériques**, Sorbonne Université, Paris (France), Faculté des Lettres, École doctorale III « Littératures françaises et comparée » (ED 019) / CELLF (Centre d'étude de la langue et des littératures françaises), UMR 8599 / OBTIC (Observatoire des textes, des idées et des corpus).

Intitulé de thèse : [DANS LES PETITS PAPIERS DE CHARCOT : DE L'EXPÉRIMENTATION AUX PRÉMISSES DE LA NEUROLOGIE MODERNE](#).

**Résumé** : Ce projet de thèse, à la jonction de l'histoire des sciences et des Lettres, propose une étude pluridisciplinaire dont l'objectif est la valorisation du fonds patrimonial de Jean-Martin Charcot, fondateur de la neurologie moderne au XIX<sup>e</sup> siècle en France. La première contribution de ce travail est la constitution d'un corpus numérique à partir des archives en question déjà numérisées. Comme deuxième contribution, nous proposons la mise en place d'une chaîne de traitement semi-automatique consistant en l'océrisation, la correction de sortie OCR, la structuration des données au format XML-TEI, la fouille sémantique et l'alignement des textes pour étudier le transfert interdisciplinaire du discours médical de Charcot dans les écrits réalisés en collaboration et dans ceux de ses continuateurs et disciples. Ces traitements nous permettront de produire une transcription interrogeable dans notre cadre de recherche. Au-delà des finalités de ce projet de thèse, ce modèle généralisable sera aussi applicable à d'autres projets de numérisation et de valorisation des fonds patrimoniaux.

2019 – 2024 **Certificat de spécialisation en linguistique**, Université de Genève (Suisse), Faculté des Lettres, Département de linguistique.

Intitulé du mémoire soutenu : [ANALYSE DES CHAMPS LEXICAUX DANS LE FONDS PATRIMONIAL DE JEAN-MARTIN CHARCOT](#).

2017 – 2019 **Master « Informatique pour les sciences humaines »**, Université de Belgrade (Serbie), programme d'études pluridisciplinaires.

Intitulé du mémoire soutenu : [CRÉATION ET ANALYSE DU CORPUS DE TEXTES DE CHANSONS ROCK YUGOSLAVES DE 1945 À 2003](#).

**Résumé** : The thesis analyzes, from the theoretical and practical perspective, the creation and processing of corpus of rock songs' lyrics originating from the former Yugoslavia in the period 1945-2003. The lyrics are obtained from the LyricWiki website using the Python library `lyricsmaster` in the *web scraping* process. The collected texts are then merged into a single XML file and automatically annotated with the `yattag` Python tool. Afterwards, the data preprocessing was conducted at the formal and content level. Furthermore, the XML document is transformed into XHTML format applying XSLT processor, in order to generate basic corpus data. The diacritic restoration process with the "Slovo Majstor" application and morphological electronic dictionaries of Serbian language in the LeXimir software package, is also automated. The *text mining* process encompassed retrieving socio-political and patriotic topics using `NLTK` library in Python, while romantic and other topics were visualized using the TreeCloud and WordItOut software. The similarity between authors represented in the corpus was measured using `stylo` package in the programming language R. Finally, an overview of the today's most relevant programming libraries in the field of *natural language processing* is provided, which, at the same time, serves as a guideline for the future work.

**Mots clés** : corpus linguistics, Yugoslav rock and roll, web scraping, natural language processing, text mining.

- 2017 – 2018 **Master « Langue, littérature et culture », Université de Belgrade (Serbie), Faculté de Philologie, module Langue, littérature et culture.**  
 Intitulé du mémoire soutenu : **L’UTILISATION DE LA PLATEFORME MOODLE DANS LES ÉTUDES NÉO-HELLÉNIQUES RÉFORMÉES.**  
**Résumé :** This paper deals with the Modern Greek language and literature learning using the Moodle electronic platform within the *hybrid* and *online* learning model. The subject of the discussion are the main features of this teaching method, as well as the application of Moodle in working with the first-year undergraduate students of the Department of Neohellenistic Studies at the Faculty of Philology, University of Belgrade. We examine the effectiveness of the mixed-mode learning by presenting the electronic contents that are available on the online courses *Contemporary Greek Language G-1* and *G-2*, *Greek Literature 1* and *Practicum in Neohellenistics 2*. Special attention is devoted to the analysis of the course interactivity, multimediality, and flexibility. Finally, we also expose the results of the accompanying empirical research, which led to the conclusion that the implementation of Moodle in teaching significantly helps students in learning.  
**Mots clés :** hybrid learning, Moodle, modern Greek language, literature, university teaching.
- 2013 – 2017 **Licence « Langue, littérature et culture grecques modernes », Université de Belgrade (Serbie), Faculté de Philologie, Département d’études néo-helléniques.**
- 2009 – 2013 **Lycée philologique (option français), Lycée philologique de Belgrade (Serbie).**

## Expérience professionnelle

Les volumes sont renseignés en heures équivalent TD (HETD).

- 2024-2026 ATER, Sorbonne Université, UFR Sociologie et Informatique pour les Sciences Humaines, Licence Sciences du langage et Master Langue et Informatique.

Libellé ELP	Code ELP	Public	Année	Volume
PIX – Formation aux compétences numériques	PIX@S2N2	L1 / L2 / L3	2024-2026	TD : 127,5 CM : 3
Analyse de discours avec programmation	L2SOPROG	L1	2024-2026	TD : 2
Corpus, ressources et linguistique outillée	M2SOL034	M1	2024-2026	CM : 21 TD : 21
Statistiques inférentielles	L6SOSTAT	L3	2024-2026	TD : 36,5
Recherche d’information et site web	L3SOWEB1	L2	2024-2025	TD : 78
Mathématiques	L5SOMATH	L3	2024-2025	TD : 9
<b>TOTAL : 298</b>				

- 2024 AGENT TEMPORAIRE VACATAIRE (ATV), cours « [Approfondissements – Corpus numériques \(H5MUAB7\)](#) », M2 Archives · M2 Sciences de l’information et des bibliothèques (SIB), Université d’Angers (France), Faculté des lettres, langues et sciences humaines (LLSH), 3h TD.
- 2023-2024 CHARGÉE D’ENSEIGNEMENT VACATAIRE (CEV), Sorbonne Nouvelle, Paris (France), UFR Littérature, linguistique, didactique (LLD) | Institut de Linguistique et Phonétique Générales et Appliquées (ILPGA), mineure Humanités numériques.

Libellé ELP	Code ELP	Public	Année	Volume
Introduction aux humanités numériques	L1HN001	L1	2023-2024	TD : 26
Cultures numériques avancées	L2HN001	L1	2023-2024	TD : 26
<b>TOTAL : 52</b>				

- 2021 – DOCTORANTE en humanités numériques, Sorbonne Université, Paris (France), Faculté des Lettres, École doctorale III « Littératures françaises et comparée » (ED 019) | CELLF (Centre d'étude de la langue et des littératures françaises), UMR 8599 | OBTIC (Observatoire des textes, des idées et des corpus), sous la direction de Prof. Dr Glenn Roe et le co-encadrement de Dr Motasem Alrahabi [[thèse](#)].
- 2021 – 2023 INTERVENANTE EXTERNE dans le cadre du projet [IDEA](#) pour la tâche d'annotation linguistique, Faculté de psychologie et des sciences de l'éducation, université de Genève (Suisse), sous la direction de Dr Joanna Blochowiak.
- 2021 MEMBRE DU COMITÉ du programme des actes de la conférence [RANLP 2021](#). Évaluation de l'article Rozovskaya, A. (2021). “[Spelling Correction for Russian : A Comparative Study of Datasets and Methods](#)”. *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*.
- 2021 STAGIAIRE au sein de l'Observatoire de la Vie Littéraire (labex OBVIL), dans le cadre du projet de la Très Grande Bibliothèque ([TGB](#)) visant à améliorer la qualité de l'OCR dans la TGB pour la tâche de la reconnaissance d'entités nommées (REN), Sorbonne Université, Paris (France), sous la direction de Prof. Dr Glenn Roe et de Dr Motasem Alrahabi.
- 2021 OPÉRATRICE DE SAISIE de sources pour le projet FNS [Visual Contagions](#) (catalogues d'expositions, OCR-XML-TEI-CSV) [[GitLab](#)], Faculté des Lettres, Chaire des Humanités numériques, Université de Genève (Suisse), sous la direction de Prof. Dr Béatrice Joyeux-Prunel.
- 2020 COLLABORATRICE SCIENTIFIQUE et INGÉNIEURE DE RECHERCHE au sein du projet FNS [Manuscripts for Sale : Automatic Structuration and Annotation of Catalogues with Complex Layouts \(Katabase\)](#), Faculté des Lettres et Sciences Humaines, Institut de littérature française, Université de Neuchâtel : OCR (Transkribus et Kraken) ; correction manuelle de l'OCR des catalogues de vente de manuscrits du XIX<sup>e</sup> s. ; encodage sémantique semi-automatique et restructuration des données en XML-TEI générées par *GROBID-dictionaries* pour le remplissage de la base de données dédiée, sous la direction de Dr Simon Gabay.
- 2019 MEMBRE DU COMITÉ du programme des actes de la conférence [RANLP 2019](#). Évaluation de l'article Zosa, E., & Granroth-Wilding, M. (2019). “[Multilingual dynamic topic model](#)”. *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*.
- 2017 – 2018 TRADUCTRICE dans le projet de traduction portugais-serbe des œuvres littéraires lusophones, Faculté de Philologie, Institut Camões, Université de Belgrade (Serbie), coordinatrice : Dr Anamarija Marinović.
- 2016 – 2018 TRADUCTRICE des contrats financiers en anglais/serbe/grec moderne, Faculté de Philologie, Département d'études néo-helléniques, Université de Belgrade (Serbie), coordinatrice : Dr Anka Radenović.
- 2015 – 2016 ENSEIGNANTE des cours particuliers de grec moderne (niveau lycée).
- 2015 – 2016 RÉDACTRICE du dictionnaire grec-serbe adapté aux étudiants du Département d'études néo-helléniques dans le projet étudiant en lexicographie, Faculté de Philologie, Université de Belgrade (Serbie), coordinateur : Dr Vojkan Stojičić.

## Liste des publications

### Revues avec comité de lecture

- 2024 Petkovic, Lj., Koudoro-Parfait, C., Desmarest, M.-S., & Lejeune, G. (2024). Quelle solution pour améliorer les performances de la reconnaissance d'entités nommées sur des données bruitées, corriger l'entrée ou filtrer la sortie ?. *Revue Corpus 26 – 2025*. DOI : <https://doi.org/10.4000/1364s>

- 2024 Koudoro-Parfait, C., Petkovic, Lj., & Roe, G. (2024). Analyse multilingue de l'impact de la correction automatique de la ROC sur la reconnaissance d'entités nommées spatiales dans des corpus littéraires. *Revue TAL : Robustesse et limites des modèles de traitement automatique des langues*, 64(2). : [https://www.atala.org/sites/default/files/TAL\\_64\\_2\\_2.pdf](https://www.atala.org/sites/default/files/TAL_64_2_2.pdf)
- 2022 Petkovic, Lj., Alrahabi, M., & Roe, G. (2022). Impact de la correction automatique de l'OCR/HTR sur la tâche de reconnaissance d'entités nommées dans un corpus bruité. *Journal of Information Sciences* 21(2), pp. 1-18. DOI : [10.34874/IMIST.PRSM/jis-v21i2.36599](https://doi.org/10.34874/IMIST.PRSM/jis-v21i2.36599)
- 2019 Petković, Lj. (2019). Creation and Analysis of the Yugoslav Rock Song Lyrics Corpus from 1967 to 2003. *Infotheca – Journal for Digital Humanities*, 19(1), pp. 5-29. <[hal-03091121](https://hal.archives-ouvertes.fr/hal-03091121)>
- 2018 Petković, Lj. (2018). Upotreba onlajn platforme Mudl (Moodle) u reformisanoj nastavi Katedre za neohelenske studije Filološkog fakulteta Univerziteta u Beogradu (Utilisation de la plateforme en ligne Moodle dans l'enseignement réformé du Département d'études néo-helléniques de la Faculté de Philologie à l'Université de Belgrade). *Methodical Perspectives*, 9(9), pp. 179-200. <[hal-03091176](https://hal.archives-ouvertes.fr/hal-03091176)>

### Colloques avec comité de lecture

- 2023 Petkovic, Lj., Alrahabi, M., & Roe, G. (2023). Circulation du discours médical de Jean-Martin Charcot : premières observations. *Humanistica 2023*. Association francophone des humanités numériques. <[hal-04107099](https://hal.archives-ouvertes.fr/hal-04107099)>
- 2020 Gabay, S., Petkovic, Lj., Bartz, A., Levenson, M. G., & du Noyer, L. R. (2020). Katabase : À la recherche des manuscrits vendus. *Humanistica 2021*. Association francophone des humanités numériques. <[hal-03066108](https://hal.archives-ouvertes.fr/hal-03066108)>

### Journées d'étude avec comité de lecture

- 2024 Petkovic, Lj., Koudoro-Parfait, C., & Lejeune, G. (2024). Mes données ou tes méthodes ? Comment trancher la responsabilité côté TAL ou côté HN quand les résultats ne sont pas au rendez-vous ?. Poster présenté à la journée d'étude du GDR TAL CNRS « Traitement Automatique des Langues et les Humanités Numériques », le 7 novembre 2024, université La Rochelle.
- 2022 Petkovic, Lj., Alrahabi, M., & Roe, G. (2022). Impact de la correction automatique de l'OCR/HTR sur la tâche de reconnaissance d'entités nommées dans un corpus bruité. *Actes de la journée d'étude sur la robustesse des systèmes de TAL*, p. 22. HAL : <[hal-03853541v1](https://hal.archives-ouvertes.fr/hal-03853541v1)>

### Congrès avec comité de lecture

- 2023 Alrahabi, M., Cordova, J., Fedchenko, V., Petkovic, Lj., & Roe, G. (2023). Pandore : a toolbox for digital humanities text-based workflows. *Digital Humanities 2023 : Book of Abstracts*. Digital Humanities 2023 · Collaboration as Opportunity (DH2023). DOI : [10.5281/zenodo.8107733](https://doi.org/10.5281/zenodo.8107733)
- 2018 Stojičić, V., et Petković, Lj. (2018). Od oglasa do platforme Moodle – primer dobre prakse u nastavi modernog grčkog kao stranog (D'une publicité à la plateforme Moodle – un exemple de bonne pratique dans l'enseignement du grec moderne comme langue étrangère). *The 6<sup>th</sup> International Congress “Applied Linguistics Today” – Language, Literature and Interdisciplinarity : Book of Abstracts*. Faculté de Philologie, Université de Belgrade (Serbie). [résumé]

### Conférences avec comité de lecture

- 2022 Cordova, J. M., Dupont, Y., Petkovic, Lj., Gawley, J., Alrahabi, M., & Roe, G. (2022). Toolbox : une chaîne de traitement de corpus pour les humanités numériques. *Traitement Automatique des Langues Naturelles*, pp. 11-13. <[hal-03701464](https://hal.archives-ouvertes.fr/hal-03701464)>

- 2020 Gabay, S., du Noyer, L. R., Levenson, M. G., Petkovic, Lj., & Bartz, A. (2020). Quantifying the Unknown : How many manuscripts of the marquise de Sévigné still exist ?. *Digital Humanities DH2020*. ADHO. [<hal-02898929>](https://hal.archives-ouvertes.fr/hal-02898929)
- 2018 Petković, Lj. (2018). Implementacija tehnika analize teksta u programskom jeziku Python pomoću NLTK-a (Implémentation de techniques de fouille de textes dans le langage de programmation Python en utilisant le package NLTK). *2018 26<sup>th</sup> Telecommunications Forum (TELFOR 2018)*, pp. 975-979. IEEE. [<hal-03091167>](https://hal.archives-ouvertes.fr/hal-03091167)

### **Ateliers avec comité de lecture**

- 2021 Scius Bertrand, A., Gabay, S., Petkovic, Lj., Janes, J., Corbières, C., & Clérice, T. (2021). The BIR database – Identifying typographic emphasis in list-like historical documents. *HIP@ICDAR21 — The 6<sup>th</sup> International Workshop on Historical Document Imaging and Processing*. [<hal-03355683>](https://hal.archives-ouvertes.fr/hal-03355683)

---

## **Communications**

### **Colloques**

- 2025 LE TEXTE DE L'AUTRE. DIALOGUE INTERDISCIPLINAIRE AUTOUR DE L'INTERTEXTUALITÉ ET DU DISCOURS RAPPORTÉ – « Analyse de l'impact scientifique de Jean-Martin Charcot sur la littérature médicale en français du XIX<sup>e</sup> au XX<sup>e</sup> siècle », organisé par le Centre de recherches inter-langues sur la signification en contexte (CRISCO). Sorbonne Nouvelle, Maison de la Recherche | École des Chartes, 1-3 juillet 2025, Paris (France).
- 2023 HUMANISTICA 2023 – « Circulation du discours médical de Jean-Martin Charcot : premières observations ». 4<sup>e</sup> édition du colloque Humanistica, 23-28 juin 2023, Genève (Suisse).
- 2022 CIHN'22 – « Impact de la correction automatique de l'OCR/HTR sur la tâche de reconnaissance d'entités nommées dans un corpus bruité ». 1<sup>e</sup> édition de la Conférence Internationale « Les humanités numériques dans la gestion du patrimoine culturel, des données et de l'information. Les méthodes, les dispositifs et les perspectives » (CIHN'22), 22-23 novembre 2022, Rabat (Maroc).
- 2022 UNIGE DATA SCIENCE DAY – « The evolution of a medical concept through a data science lens. The strange case of hysteria ». 3<sup>e</sup> édition du colloque UNIGE Data Science Day, « Promises of Artificial Intelligence », 15 septembre 2022, Genève (Suisse).
- 2022 TALN-RÉCITAL – « Pandore : a toolbox for digital humanities text-based workflows ». Démonstration présentée avec Johanna Cordova et Yoann Dupont. 29<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles (TALN), « Vers un TAL inclusif », 27 juin – 1 juillet 2022, Avignon (France).
- 2021 HUMANISTICA 2021 – « Katabase : À la recherche des manuscrits vendus », avec Simon Gabay, Alexandre Bartz, Matthias Gilles Levenson et Lucie Rondeau du Noyer. 2<sup>e</sup> édition du colloque Humanistica, 10-12 mai 2021, Rennes (France).
- 2020 DH2020 – « Quantifying the Unknown : How many manuscripts of the marquise de Sévigné still exist ? », avec Simon Gabay, Alexandre Bartz, Matthias Gilles Levenson et Lucie Rondeau du Noyer. 15<sup>e</sup> édition de la conférence internationale Digital Humanities DH2020, « carrefours/intersections », ADHO, 22-24 juillet 2020, Ottawa (Canada) [dépôt des diapositives uniquement en raison du COVID].
- 2018 TELFOR – « Implémentation de techniques de fouille de textes dans le langage de programmation Python en utilisant le package NLTK », 26<sup>th</sup> Telecommunications Forum, Centre Sava, Belgrade (Serbie), 26-27 novembre 2018.

## Journées d'étude

- 2025 JOURNÉE D'ÉTUDE : HUMANITÉS NUMÉRIQUES ET INTELLIGENCE ARTIFICIELLE : INTERACTIONS / ALTERCATIONS – session posters « Analyse de l'influence de Jean-Martin Charcot par extraction de terminologie médicale : approche *PatternRank* ». Journée d'étude du lancement de l'Initiative Humanités Numériques de Sorbonne Université. Maison de la recherche, Paris (France), 7 novembre 2025.
- 2025 JOURNÉE DE TRAVAIL DU CONSORTIUM ARIANE (GT4) « Analyse littéraire assistée par l'IA » – « ... ». Journée d'étude organisée par le consortium ARIANE. SCAI, Paris (France), 17 juin 2025.
- 2024 INTELLIGENCE ARTIFICIELLE ET HUMANITÉS NUMÉRIQUES – « Mesurer l'influence de Charcot sur ses contemporains à l'aide de l'extraction des mots-clés ». Journée d'étude organisée sous l'égide de l'Association Française pour l'Intelligence Artificielle (AFIA), et des GDR CNRS MADICS et MAGIS, Bibliothèque nationale de France, Paris (France), 3 mai 2024.
- 2023 BRUIT DE FOND OU VALEUR AJOUTÉE ? Gérer le bruit lors des traitements informatiques des corpus linguistiques – « Correction automatique des interférences OCR dans la reconnaissance d'entités nommées spatiales : réel gain ou perte de l'information ? », avec Caroline Koudoro-Parfait. Journée d'étude co-organisée par l'Université Grenoble Alpes et Roma La Sapienza, Maison des langues et des cultures, Université Grenoble Alpes (France), 28 avril 2023.
- 2023 JOURNÉE « HUMANITÉS NUMÉRIQUES » – Présentation de l'équipe-projet OBTIC et du projet doctoral « Dans les petits papiers de Charcot : de l'expérimentation aux prémisses de la neurologie moderne », avec Glenn Roe et Motasem Alrahabi. 1<sup>e</sup> journée d'étude sur les humanités numériques organisée par Sorbonne Université, Maison de la recherche, Paris (France), 18 janvier 2023.
- 2022 ROBUSTAL22 – « Améliorer la qualité de l'OCR dans la TGB pour la tâche de REN ». 1<sup>e</sup> journée d'étude sur la robustesse des systèmes de TAL organisée par Caio Corro et Gaël Lejeune, Maison de la recherche, Paris (France), 25 novembre 2022.

## Congrès

- 2023 DH2023 – « Pandore : a toolbox for digital humanities text-based workflows ». Poster présenté avec Glenn Roe, Motasem Alrahabi, Johanna Cordova et Valentina Fedchenko. 18<sup>e</sup> édition du congrès international “DH2023 – Collaboration as Opportunity”, Université de Graz | Messe Congress Graz, Graz (Autriche), 10–14 juillet 2023.
- 2018 ALT 6 – « Od oglasa do platforme Moodle – primer dobre prakse u nastavi modernog grčkog kao stranog » (D'une publicité à la plateforme Moodle – un exemple de bonne pratique dans l'enseignement du grec moderne comme langue étrangère), avec Vojkan Stojićić. 6<sup>e</sup> congrès international “Primenjena lingvistika danas – jezik, književnost i interdisciplinarnost” (“La linguistique appliquée aujourd’hui – langue, littérature et interdisciplinarité”), Faculté de Philologie, Belgrade (Serbie), 12–13 octobre 2018.

## Séminaires

- 2025 SÉMINAIRE DOCTORAL DE L'ÉQUIPE OBTIC. Séminaire organisé par Prof. Dr Glenn Roe et Dr Motasem Alrahabi, SCAI, Paris (France).
- « Mesurer l'impact de Jean-Martin Charcot *via* l'extraction des termes médicaux : quelle approche adopter ? », le 6 février 2025.
  - « Comparaison des approches pour l'extraction de phrases-clés », le 13 mars 2025.
  - « Extraction de phrases-clés à partir du corpus Charcot : évaluations quantitative et qualitative », le 15 avril 2025.

- 2024 SÉMINAIRE DE L'ÉQUIPE « LINGUISTIQUE COMPUTATIONNELLE – RENCONTRES MINUTE » – « Origines et progrès du projet Charcot ». Séminaire organisé par Dr Gaël Lejeune, 12 septembre 2024, Maison de la Recherche, Paris (France).
- 2024 SÉMINAIRE · ATELIER DOCTORAL SUR L'HTR ET eSCRIPTORIUM – Retour d'expérience. Séminaire organisé par le Centre Roland Mousnier de la Faculté des Lettres de Sorbonne Université, 26 avril 2024, Centre Roland Mousnier, Paris (France).
- 2024 SÉMINAIRE DOCTORAL DE L'UNITÉ DE SERVICE CERES, « APPROCHES DES MÉTHODES NUMÉRIQUES » – « Dans les petits papiers de Charcot : de l'expérimentation aux prémisses de la neurologie moderne ». Séminaire organisé par Léa Andolfi, Julien Bezançon, Clara Bordier, Thibault Grison, Rimane Karam, Adélie Laruncet et Marine Tiger, 27 mars 2024, Maison de la Recherche, Paris (France).
- 2023 SÉMINAIRE POUR LES ÉTUDIANTS EN MASTER, « LES HUMANITÉS NUMÉRIQUES LITTÉRAIRES » – « Circulation du discours médical de Jean-Martin Charcot : premières observations ». Séminaire organisé par Prof. Dr Glenn Roe, 6 décembre 2023, Maison de la Recherche, Paris (France).
- 2023 SÉMINAIRE · ATELIER DOCTORAL SUR L'OCR ET L'HTR – « Numérisation de documents : transcription, annotation, OCR, HTR, correction... », avec Dr Motasem Alrahabi. Séminaire organisé par le Centre Roland Mousnier de la Faculté des Lettres de Sorbonne Université, 22 novembre 2023, Maison de la Recherche, Paris (France).
- 2023 SÉMINAIRE DE L'ÉQUIPE « LINGUISTIQUE COMPUTATIONNELLE » – « Correction automatique des interférences OCR dans la reconnaissance d'entités nommées spatiales : réel gain ou perte de l'information ? », avec Caroline Koudoro-Parfait. Séminaire organisé par Dr Gaël Lejeune, 13 avril 2023, Maison de la Recherche, Paris (France).
- 2022 SÉMINAIRE POUR LES ÉTUDIANTS EN MASTER, « LES HUMANITÉS NUMÉRIQUES LITTÉRAIRES », « Impact de la correction automatique de l'OCR/HTR sur la tâche de reconnaissance d'entités nommées dans un corpus bruité ». Séminaire organisé par Prof. Dr Glenn Roe, « Impact de la correction automatique de l'OCR/HTR sur la tâche de reconnaissance d'entités nommées dans un corpus bruité », 16 novembre 2022, Maison de la Recherche, Paris (France).
- 2021 SÉMINAIRE JERTEH – « Katabase : In Search of Lost Manuscripts », avec Simon Gabay et Alexandre Bartz. Séminaire organisé par JERTEH, Société des ressources et des technologies linguistiques à Belgrade (Društvo za Jezičke Resurse i Tehnologije, Beograd), 10 juin 2021, Belgrade (Serbie).

## Ateliers

- 2024 ATELIER PYTHON – « Extraction de concepts-clés à partir du fonds Charcot ». Atelier organisé par l'équipe-projet OBTIC, 30 avril 2024, DataLab, Bibliothèque nationale de France, Paris (France).
- 2023 ATELIER HTR – « Utilisation des techniques de transcription automatique des textes pour l'analyse de documents maritimes dactylographiés du XX<sup>e</sup> s. », avec Caroline Koudoro-Parfait. École d'été d'Analyse Numérique de l'HIstoire de la Mer et des Océans (ANHIMO 2023), organisée par l'institut de l'Océan de l'Alliance Sorbonne Université et Sorbonne Centre for Artificial Intelligence (SCAI), Paris (France), 26-30 juin 2023.
- 2021 ATELIER PYTHON – « Manipuler les fichiers texte ». Atelier organisé par l'équipe-projet OBTIC, 5 décembre 2021, SCAI, Paris (France).
- 2021 ATELIER OCR – « Correction automatique des sorties d'OCR ». Atelier organisé par l'équipe-projet OBTIC, 25 novembre 2021, SCAI, Paris (France).
- 2021 ATELIER OCR – « Transkribus, Kraken, eScriptorium, Tesseract », avec Johanna Cordova. Atelier organisé par l'équipe-projet OBTIC, 18 novembre 2021, SCAI, Paris (France).

- 2021 ATELIER OCR – « Outils de numérisation de documents : Transcription, OCR, HTR... », avec Motasem Alrahabi. Atelier organisé par l'équipe-projet OBТИC, 28 octobre 2021, SCAI, Paris (France).

### Posters

- 2025 JOURNÉE D'ÉTUDE : HUMANITÉS NUMÉRIQUES ET INTELLIGENCE ARTIFICIELLE : INTERACTIONS / ALTERCATIONS – session posters « Analyse de l'influence de Jean-Martin Charcot par extraction de terminologie médicale : approche *PatternRank* ». Journée d'étude du lancement de l'Initiative Humanités Numériques de Sorbonne Université. Maison de la recherche, Paris (France), 7 novembre 2025.
- 2024 JOURNÉE D'ÉTUDE GDR TAL CNRS « TRAITEMENT AUTOMATIQUE DES LANGUES ET LES HUMANITÉS NUMÉRIQUES » – Mes données ou tes méthodes ? Comment trancher la responsabilité côté TAL ou côté HN quand les résultats ne sont pas au rendez-vous ?. Poster présenté avec Gaël Lejeune, le 7 novembre 2024, université La Rochelle.

### Co-encadrement de mémoires de master

- 2025 Camille Fief et Yeleen Canonge (M1 – Sorbonne Université) : « Reconnaissance d'entités nommées et entraînement de modèle sur données bruitées », avec Gaël Lejeune et Caroline Koudoro-Parfait.
- 2025 Ruixing Zheng et Marina Carvalho Lopes (M1 – Sorbonne Université) : « Reconnaissance d'Entités Nommées sur un corpus de poèmes du 19<sup>e</sup> siècle », avec Gaël Lejeune et Caroline Koudoro-Parfait.

### Formations complémentaires

- 2025 [DFC] ETHIQUE DE LA RECHERCHE ET INTÉGRITÉ SCIENTIFIQUE, organisée par Sorbonne Université, campus Cordeliers, Sorbonne Université, Paris (France) [participation : présentiel] [attestation de présence].
- 2023 SUMMER SCHOOL ON DIGITAL HUMANITIES – école d'été en humanités numériques organisée dans le cadre de l'alliance stratégique entre Université nationale autonome du Mexique (UNAM) et Sorbonne Center for Artificial Intelligence (SCAI). Institut de recherches en mathématiques appliquées et en systèmes (IIMAS), Ville de Mexico, Mexico [attestation de présence].
- 2022 ENEXDI (*Encoder, Exploiter, Diffuser*) – « Les humanités numériques dans les projets de recherche », école de formation intensive en humanités numériques organisée par la Maison des Sciences de l'Homme et de la Société (MSHS), Université de Poitiers, France [attestation de présence].
- 2022 « INTRODUCTION AU TRAITEMENT AUTOMATIQUE DES LANGUES » – cours proposé dans le cadre du Master « Humanités numériques » de l'université PSL (Paris Sciences & Lettres) | ENC (École nationale des chartes) [attestation de présence].

### Bourses et prix

- 2023 – BOURSE DE DÉPLACEMENT ADHO de la conférence DH2023 à Graz (Autriche).
- 2021 – FINANCEMENT DU CONTRAT DOCTORAL PAR L'INITIATIVE D'EXCELLENCE (IDEX) dans le cadre des Instituts et des Initiatives de l'Alliance Sorbonne Université (ASU) (institut Observatoire des Patrimoines de l'Alliance Sorbonne Université – OPUS).
- 2017 – 2018 DOSITEJA – Bourse du gouvernement serbe (Fondation pour les jeunes talents) pour les meilleurs étudiants en master.
- 2016 – 2017 DOSITEJA – Bourse du gouvernement serbe (Fondation pour les jeunes talents) pour les meilleurs étudiants en licence.

- 2016 **MENTION DE LA DOYENNE** pour l'excellence académique, Faculté de Philologie, Université de Belgrade (Serbie).
- 2014 – 2016 **BOURSE DU GOUVERNEMENT SERBE** du Ministère de l'éducation, des sciences et du développement technologique pour les étudiants en licence.
- 2014 **CERTIFICAT DE FORMATION** du Centre de formation et d'apprentissage tout au long de la vie de l'Université d'Ioannina (Centre d'enseignement et de culture de la langue grecque DI.K.E.P.P.E.E. "Stavros Niarchos") – bourse IKY pour l'École d'été de la langue et culture grecques à Ioannina (Grèce).

## Compétences

### Informatiques

MS Office Word, Excel, Access, PowerPoint

SE Windows, macOS, Linux

### Langages

Script Python (`os, re, numpy, matplotlib, scipy, pandas, scikit-learn, nltk, tensorflow, pytorch, gensim, requests, spacy`...), R (`stylo, ggplot2, knitr`), Bash, Perl

Développement HTML/CSS

Base de données SQL

Balisage XML, L<sup>A</sup>T<sub>E</sub>X, Markdown

Feuilles de style XSLT

### Gestion de versions

Git

### Autres logiciels

Transkribus, Kraken, eScriotorium, TXM, GROBID-dictionaries, Gephi, Docker, Oxygen XML Editor, Anaconda, Visual Studio Code, RenderX XEP Engine, DARIAH-DE Topics Explorer, SPSS, SQL Server Management Studio, Azure Data Studio, Unitex, PyCharm, TreeCloud, Zotero, PIX, Overleaf, Texmaker, TeXstudio

### Familière avec

DTS, IIIF, SPARQL, Flask, Protégé, MATLAB, Isabelle

### Audiovisuel

Final Cut Pro X, Sony Vegas, FL Studio, MAGIX, Audacity, Virtual DJ

### Linguistiques

Serbe lu, écrit, parlé : langue maternelle

Français lu, écrit, parlé : courant

Anglais lu, écrit, parlé : courant

Grec moderne lu, écrit, parlé : courant

Portugais lu, écrit, parlé : intermédiaire

Coréen lu, écrit, parlé : débutant

Latin lu, écrit

Grec ancien lu, écrit

## Expérience sportive

- 2024 – taekwondo – ceinture violette (XX *keup*), club SP Training, Bourg-la-Reine (France).
- 2021 – 2023 taekwondo – ceinture bleue-deux barrettes rouges (4<sup>e</sup> *keup*), club Paris Taekwondo Elite, Paris (France).
- 2019 – 2021 taekwondo – ceinture verte-barrette bleue (5<sup>e</sup> *keup*), club Il Gi Dojang, Genève (Suisse).
- 2015 – 2019 taiji quan – club “Peng”, Belgrade (Serbie).
- 2015 – 2019 qigong – certificat de participation aux modules I et II pour l'instructeur du qigong, Fédération serbe de qigong, Belgrade (Serbie).

## Autres intérêts

2018 – 2019 Membre de la Société des ressources et des technologies linguistiques à Belgrade (Serbie) JERTEH (Društvo za Jezičke Resurse i Tehnologije, Beograd [[lien](#)])

- Arts audiovisuels
- Langues et cultures de l'Asie de l'Est (chinoise, japonaise, coréenne)